PJM LOAD/ENERGY FORECASTING MODEL

White Paper

Capacity Adequacy Planning Department

Last Updated: February 22, 2007

Table of Contents

INTRODUCTION

This document discusses the impetus for, and development of, a load forecasting function within the PJM Interconnection. It is intended to serve as documentation of the implemented peak and energy forecast models as well as of other methods and specifications which were tested but not used. Its intended audience is members of the PJM Load Analysis Subcommittee and the Planning Committee.

BACKGROUND OF PJM ASSUMING FORECASTING ROLE

Historically, the PJM Interconnection has produced and published an annual peak load and energy forecast. The forecast, known as the PJM Load Forecast Report, was published each February and covered a ten year forecast horizon. It was used to support the Installed Reserve Margin Study, the Regional Transmission Expansion Plan (RTEP) and for reporting to the North American Electric Reliability Council and various governmental agencies.

The PJM Load Forecast Report included estimates of monthly and seasonal peak loads, monthly outputs (measured as net energy for load) and available load management. Forecasts were presented for each transmission zone in the RTO, as well as for each PJM Region (geographic area composed of aggregated transmission zones) and the RTO as a whole. The three defined Regions are the PJM Mid-Atlantic Region, PJM Western Region and PJM Southern Region. Tables of historic PJM peaks and energy demands were also included.

The forecast for each transmission zone was provided to PJM by the respective Electric Distribution Company (EDC) through its representative to the PJM Load Analysis Subcommittee (LAS). The peak load forecasts are Non-Coincident Peaks (NCP); each company projects its peaks irrespective of when the Region or RTO peaks. Each submitting entity produced its forecast based on its own methodology, though it was common that the energy forecast was derived from the company's retail sales forecast and the energy forecast was then used to derive the peak load forecast.

LAS members provided their forecasts to the PJM Capacity Adequacy Planning (CAP) Department in mid-January. CAP staff reviewed the forecasts and calculated aggregate, Coincident Peak (CP) forecasts for the Regions and RTO using a long-term average of the observed historical diversity within the Region or RTO. The adjustment for diversity reduced the sum of the zonal NCPs to reflect that zones may peak on different days and hours. Energy forecasts for Regions and the RTO were calculated as the sum of the zonal energy forecasts. CAP staff produced a draft Load Report with an executive summary, graphs and tables. The draft report was reviewed by the LAS at its February meeting. After any revisions were made, the report was submitted to the PJM Planning Committee for review and approval. Once the report was approved by the Planning Committee, it was released to the public. Prior to 1999, hard copies were sent to PJM staff and committee members. Since then, the report has been posted on the PJM.com website. The posting includes a spreadsheet file with the forecast data.

Since 2000, the LAS has produced a supplement to the Load Forecast Report, which includes projections of weekly peaks for the Regions (derived from the Regional monthly peaks) and extreme weather forecasts of the Regional peak forecasts. The extreme weather forecasts are calculated to represent the loads that would result from much warmer and much cooler than expected weather conditions. This type of weather is expected to occur once every ten years. The supplemental report was released after being reviewed by the LAS at its April meeting.

With the advent of retail choice in some jurisdictions in 1999, several factors led to the decision to develop an independent PJM load forecast to replace the diversified sum of zones forecast.

PJM's Reliability Assurance Agreements (RAAs) specify that capacity obligations for Load Serving Entities are to be based on allocated shares of forecasted zonal peaks for a planning period. When the PJM capacity market began in 1999 there was no independent forecast of peak load and PJM had no internal forecasting function. The RAA Reliability Committee therefore decided that weather normalized peak loads from the prior summer would be used in place of forecasted peaks. The weather normalized peak loads were produced by PJM staff. This approach, however, had the drawback of failing to recognize the load growth that had occurred since the previous summer.

Therefore, in 2003, PJM implemented forecasted peak loads in the capacity obligation process by using a PJM-produced RTO forecast to scale the normalized peaks from the prior year up to the expected peak of the current planning period. This forecast, dubbed Entity, used a single-equation model to estimate PJM load growth as a function of growth in the U.S. economy. As part of the transition, PJM committed itself to developing a more robust forecast model to replace Entity.

Despite the development of the Entity forecast, the diversified sum of zones forecast continued to be used to support reliability and transmission planning studies. Citing the organizational structure of an independent RTO, some market participants, state officials and stakeholders began to question the propriety of an RTO forecast linked directly to the EDCs. Additionally, some questioned the accuracy of the forecasts. Several times since 2000, forecasted peak load growth rates for the first forecasted year for some zones showed negative load growth and long-term growth rates that were inconsistent with their historical zonal growth rates. A consequence of this was that the diversified sum of zones often resulted in an RTO CP forecast that was inconsistent with historical RTO peak load growth rates. This resulted in a delay in producing an official PJM RTO forecast for use in planning studies while concerns were investigated. The Planning

Committee instructed the LAS to address these issues. Actions were taken that improved, but did not solve, the situation.

Integration of new members into the RTO added further impetus to developing an internal PJM forecasting function. As the PJM footprint grew, it became increasingly difficult to coordinate forecast inputs from external parties to create a consistent RTO forecast. As the breadth of the forecast for the PJM RTO increased, and as the forecast needed to be explained and supported in many more venues and jurisdictions, the inability of the PJM staff to respond to inquiries and display an independent assessment of inputs and results became more apparent.

In 2004, the combined impact of these influences led PJM to launch an effort to develop its own load forecasting capability.

TIMELINE OF PJM LOAD FORECAST MODEL DEVELOPMENT

Selection of Consultant - October 2004

PJM issued a Request For Proposal (RFP) to obtain additional industry expertise in developing a PJM forecasting process that would produce an independent, but consistent, forecast for the RTO and EDC zones.

The RFP requirements were:

1. CAP Staff requests assistance in developing a forecast methodology that will produce a PJM RTO CP forecast independently of the EDCs' NCP forecasts.

2. The PJM RTO CP forecast should be consistent with the sum of the EDCs' NCP forecasts.

3. The methodology should recognize, evaluate and incorporate the major drivers of annual peak electrical demand within the PJM RTO and its Regions (Mid-Atlantic, Southern and Western). These RTO and regional drivers should have readily available and consistent forecasts.

4. The sum of NCP forecasts for the three regions must be consistent with the PJM RTO CP forecast.

5. The methodology should be tested for applicability on a zone, sub-zone or cluster (combination of zones or sub-zones) level.

6. The methodology should achieve a balance between accuracy and ease of use.

7. The methodology should be able to incorporate expected (50^{th} percentile) or extreme (90^{th} percentile/10^{th} percentile) forecasts of the drivers.

8. The accuracy of the methodology should be demonstrated through a back-forecasting process.

9. Since weather is a major driver of electrical demand, the methodology should address the issue of weather-normalization of data.

10. The methodology will be developed in close collaboration with CAP staff.

After competitive bidding and a formal review process, PJM selected Itron Inc. to work on the project. In addition to extensive industry expertise, Itron brought a robust modeling platform (MetrixND, SQL Server) that would allow PJM to easily create, update, and refine the forecast models on a timely basis. In addition, Itron had a methodology to simulate weather scenarios across a broad geographic area.

Initial Model Development With Itron - December 2004

CAP staff began work with Itron to develop zonal NCP and CP forecast models. The project kicked off with PJM visiting Itron's facilities for a 3-day intensive model development session. PJM provided Itron with a general idea of the concepts to be incorporated in the forecast system: load, weather, economic, and temporal effects.

Itron presented four potential model structures to forecast daily NCP loads:

1. a multiple regression specification using county-level economic drivers
2. a multiple regression specification using U.S. Gross Domestic Product
3. a neural network specification using county-level economic drivers
4. a neural network specification using U.S. Gross Domestic Product

Given the long-term (10-year) focus of the PJM forecast, PJM concluded it would be more appropriate to use an econometric regression framework for both the NCP and CP models, rather than to use a neural network process. Although the neural network model may be more robust in estimating the nonlinear relationships inherent in system demand, the econometric regression framework allows for easier interpretation and understanding of model coefficients and other statistics. Econometric models also have the advantage of being more compatible with use in scenario analysis.

CAP staff also concluded at the time that using U.S. Gross Domestic Product (GDP) would be preferable to using county-level economic drivers for two reasons: vintage (delay in reporting history) of county-level data and ease of acquisition. While stakeholders would have to contract with an economic data provider to replicate county-level forecast assumptions, U.S. GDP is more easily obtained, widely available and updated on a more timely basis.

At the end of the initial model development stage, Itron, in conjunction with PJM, specified and estimated 18 zonal regression models to create both NCP and CP forecasts. The regression models related zonal loads to changes in the U.S. GDP, calendar and weather. The models were evaluated based on their in-sample (historical) fit. At this point, the weather simulation process had not been completed.

Forecast Modeling Framework

There are several dimensions to developing peak forecasts for a control region as geographically diverse as the PJM RTO. The forecasting framework must account for temporal, weather, and economic diversity.

- **Temporal Diversity**. As PJM grew, there was a strong possibility that the zones would not peak at the same day or hour as the PJM RTO. For example, a heat wave in the Midwest will impact the PJM Western Region, but not necessarily the Mid-Atlantic and Southern Regions. As a result of the heat wave, the zones in the Western Region could very well peak on a day different from the day of the PJM RTO peak. Also, the PJM RTO spans the Eastern and Central time zones. The forecasting framework must be sufficiently flexible to allow each zone's NCP to occur on a date and time combination that is not necessarily the day of the PJM RTO peak.
- **Weather Diversity**. Because of the size of the PJM RTO, the prevailing weather could be vastly different across the region. Traditionally, utilities have used "normal" or "extreme" weather conditions to drive their long-term forecasts. When thinking of a zone in isolation from the PJM RTO, this continues to be a viable approach to account for the impact of weather on peak loads. In the context of the PJM RTO, the idea of "normal" or "extreme" weather becomes problematic. Ideally, the forecasting framework should allow long-term peak forecasts to be generated under realistic weather patterns that allow for temporal diversity. At the same time, the framework needs to support construction of normal (50^{th} percentile), extreme (90^{th} percentile) and mild (10^{th} percentile) peak forecasts.
- **Economic Diversity**. Economic diversity has two dimensions. First, over a ten year forecast horizon there could be significant load growth within a zone due to additional homes and increased business activity. General growth in the number of homes and businesses in a region will lead to peak load growth independent of weather simply because there is more equipment installed and operating. Second, more homes and businesses mean a greater amount of air conditioning and space heating equipment. The increased saturation of weather-sensitive equipment will lead to an increase in the weather sensitivity of a zone's peak load. The forecasting framework has to allow for a differing impact of economic growth on base and weather-sensitive loads. In addition, the framework should allow these impacts to vary across zones.

Monte Carlo Framework

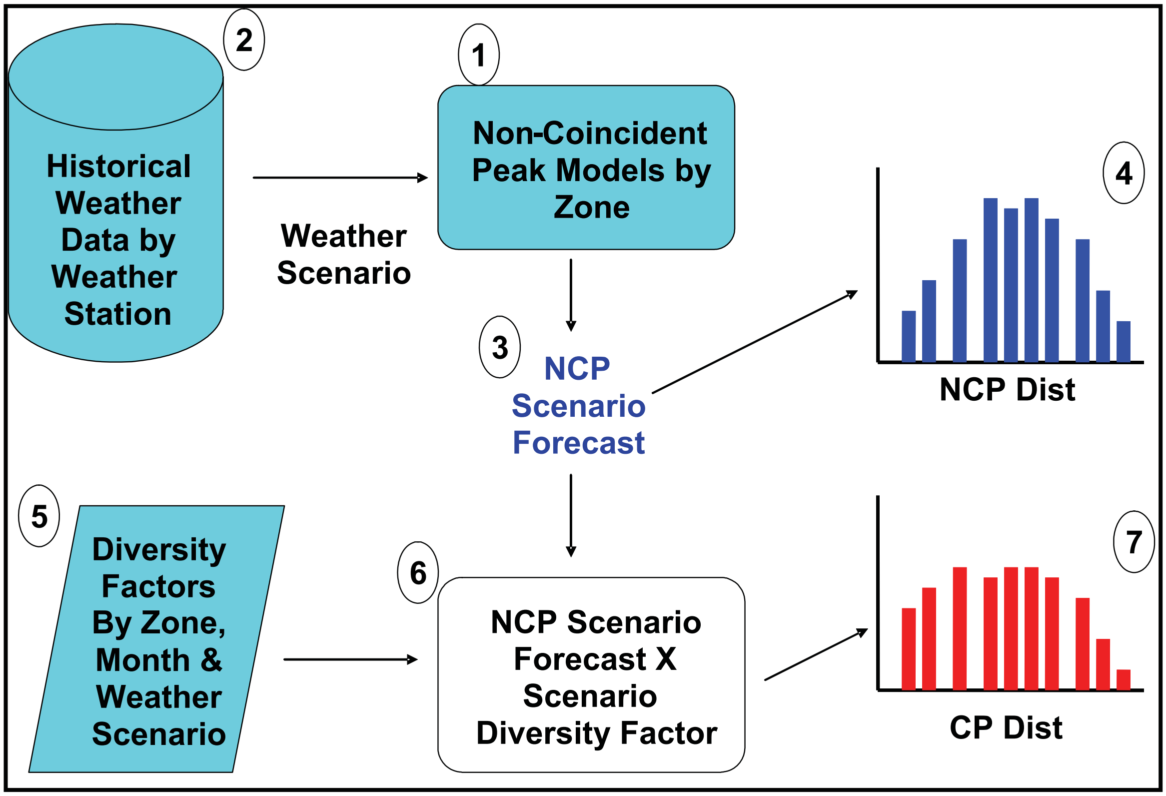
To meet these challenges, CAP implemented a Monte Carlo framework for generating Non-Coincident and Coincident Peak forecasts by zone and for the PJM RTO. Monte Carlo simulation allows for the estimation of all probable outcomes, as opposed to a single forecast value. Using this approach, load forecasts are developed for each zone using weather patterns that prevailed across the PJM RTO since 1972. In this way, the forecasts reflect how the PJM system can be expected to perform, given weather conditions that have actually occurred. In so doing, the framework addresses the issue posed by temporal diversity, by allowing the Zone forecasts to be driven by the range of actual weather conditions. This allows the NCP forecasts the flexibility to occur on different days. Further, the Monte Carlo simulations result in distributions of forecasted NCP and CP loads. The 10^{th} /50^{th} /90^{th} percentile load values can be pulled directly from these distributions. The economic diversity is addressed by incorporating economic drivers into the forecasting models, using zone-specific variables for economic growth and their interaction with weather.

The key elements of the Monte Carlo framework are depicted in Figure 1 and are described briefly below.

1. **Non-Coincident Peak Models by Zone**. Models of the daily NCP are constructed for each zone. These models are used to generate daily NCP forecasts by weather scenario. These models account for the impact of day-of-the-week, holiday, solar, weather and economic conditions on daily peak demand.
2. **Historical Weather Database**. PJM compiled an historical database of hourly weather data for 13 unique weather stations across the PJM RTO footprint. The data covers the years 1971 through the present. These actual weather patterns form the basis of the Monte Carlo simulations.
3. **NCP Scenario Forecast**. The historical weather patterns are combined with zonal NCP models to generate a set of NCP scenario forecasts.
4. **NCP Forecast Distribution**. The NCP scenario forecasts are used to construct a distribution of NCP forecasts by zone and year, by ordering them by load magnitude. These distributions are then the basis for the 10^{th}, 50^{th}, and 90^{th} percentile NCP peak forecasts.
5. **Coincidence Factors by Zone and Scenario**. Coincidence factors are computed for each zone, weather scenario and month, based on historical diversity. Coincidence factors are the inverse of diversity factors.
6. **Compute CP Forecasts**. The NCP forecasts are combined with the Coincidence Factors to generate CP forecasts by zone and weather scenario.
7. **CP Distribution Forecast**. The resulting CP forecasts are used to construct a distribution of CP forecasts by zone and year, by ordering them by load

magnitude. These distributions are then the basis for the 10th, 50th, and 90th percentile CP peak forecasts.

Figure 1: Overview of the Monte Carlo Framework



Structure of the Model

While a separate model is estimated for each PJM zone, they all share the same structure of estimating historical daily NCP as a function of independent variables. The model as specified at the end of the three-day session can be written generally as:

$$\text{E1: } \mathrm{NCP}^{u}_{y,m,d} = \mathrm{F}\left(\mathrm{Calendar}_{y,m,d}, \mathrm{Solar}^{u}_{y,m,d}, \mathrm{Economics}^{u}_{y,m,d}, \mathrm{Weather}^{u}_{y,m,d}\right)$$

Here, the daily peak for a zone indexed by (u) in a year (y), month (m) and day (d) is modeled as a function of calendar conditions (e.g., day-of-the-week, holidays and month), solar conditions (e.g., sunrise and sunset times, and a daylight savings indicator), economic conditions (e.g., GDP) and weather conditions (e.g., Temperature Humidity Index or THI). These models were estimated using historical daily demand data from 1998 through 2004 (data for the entire RTO was available only back to 1998).

The discussion will now resume outlining the model development timeline. More detail on the final model structure is presented in the "Technical Discussion" section below.

Itron Model Development Refinement - January 2005

Itron informed PJM that having independent NCP and CP regression models was not producing consistent results. With two independent models, there was no assurance that the CP forecast would be less than or equal to the NCP forecast, a necessity by definition.

To resolve this issue, Itron proposed modeling NCP-to-CP diversity over a historical time interval and applying these ratios to the NCP forecast to create the CP forecast. PJM agreed that this approach would provide consistency between the NCP and CP forecasts, so long as the modeled diversity factors were greater than or equal to one (i.e., there is diversity between the NCP and CP).

While Itron worked to further develop the diversity factor proposal, PJM pursued acquiring and benchmarking the necessary computer resources to run the forecast model in-house.

The Forecast Model described by Equation E1 is solved by moving through the year day by day, using forecasted economic variables and the historic weather scenarios for each day. The economic scenario selected to be used for the base forecast was the "most likely economic scenario" as obtained from Moody's Economy.com.

To model the most likely weather conditions (often referred to as the expected or normal weather), a Monte Carlo technique was used to simulate a distribution of daily load scenarios generated by historical weather observations from 1972 to the present which represent the actual weather patterns that prevailed across the PJM control region.

The weather parameter used was a weighted Temperature Humidity Index (THI) calculated across a combination of weather stations and calendar days. (The use of multiple daily THIs allowed the model to recognize the impact on system load of significant heat build-up over a period of days.) To enhance the simulation process, each historical weather pattern is shifted across each day of the week, providing seven different weather scenarios for each historical year.

Thirty-three historical years of weather data and the seven different weather scenarios require solving the daily NCP model 84,315 times (365x33x7) to produce a forecast of one year for one zone. A ten year forecast of 18 zones requires solving the daily NCP model 15,176,700 times (84,315x10x18).

Also, CAP worked with PJM's ITS Department to obtain sufficient computer network resources to produce solutions in a timely manner.

Initial Presentation to Stakeholders - February 2005

PJM presented the status of the forecast development process to the LAS at their February meeting. In subsequent meetings, various enhancements to the process were discussed that reflected input from the LAS.

The weather parameter associated with each zone was based on data from a weighted combination of weather stations within that zone, as selected by PJM. PJM offered LAS members the opportunity to provide alternative weather station weightings than those used in the initial development process. To the extent that model results improved or did not differ significantly from the original weighting protocol, PJM agreed to incorporate the EDC's weightings, for consistency. The results of these analyses by PJM indicated that weights provided by the zones were as good as or better than the original protocol and therefore they were incorporated in the PJM model.

Weather Simulation and Diversity Modeling Enhancements - March 2005

Itron completed the weather simulation and diversity modeling and presented the results to PJM for review. PJM was pleased with the outcome of the diversity factor simulation process as a viable solution to the NCP to CP crossover issue.

PJM completed its initial modeling phase with Itron, developed process documentation and installed the forecasting framework on PJM servers. At this point, the formal relationship with Itron ended.

PJM staff became more familiar with using MetrixND and the simulation process delivered by Itron

PJM held a special meeting with the LAS to discuss PJM management's proposal to use the PJM forecast model to set capacity obligations for summer 2005. At this LAS meeting, PJM presented the model fit statistics, coefficients, and NCP and CP forecasts.

This topic was presented to the PJM Planning Committee a few days later. Some Planning Committee members pointed out the significant impact use of the PJM forecast could have on their capacity obligations. PJM management therefore proposed, and the Planning Committee endorsed, not using the PJM model for summer 2005 capacity obligations but to instead use a transition year "Entity" model process.

Economic Driver Enhancements- May - October 2005

PJM explored using local economic drivers to replace U.S. Gross Domestic Product (GDP). Two additional model specifications were developed, one using Gross State Product (GSP) and one using Gross Metropolitan Area Product (GMP). GSP and GMP are derived from the same source data as Gross Domestic Product, but are localized to a state or Metropolitan Statistical Area. Historical and forecast data were obtained from Moody's Economy.com and were used to replace GDP as an independent variable in the NCP models. While all specifications worked well, model fit with GSP and GMP proved to be slightly better overall than with GDP, but forecasts showed significantly slower peak growth for a number of PJM regions. PJM discussed this situation with Economy.com, and it was determined that a problem existed with the allocation of U.S.-level data to some states and metropolitan areas. When this situation was rectified, the forecasts showed peak growth consistent with historical trends. Gross Metropolitan Product was selected as the economic driver because of its performance in the model and its ability to be tailored to each zone.

Model Established for 2006 Load Forecast Report - November-December 2005

PJM re-estimated diversity factor distributions and explored numerous approaches to developing CP forecasts.

PJM finalized model specification and diversity factors for use in the PJM Load Forecast Report. All NCP models used Gross Metropolitan Area Product (GMP) and CP forecasts were modeled as zonal shares of the RTO peak.

PJM staff solved the eighteen zonal models for an eleven year forecast horizon (2006 – 2016), using an economic forecast from Economy.com dated December 2005, and weather scenarios for the years 1972 through 2004. Model results were rigorously evaluated through goodness-of-fit statistical tests and back-forecasting. The PJM CP and zonal NCP forecasts were then published in the 2006 PJM Load Forecast Report in February. The 2006 PJM Load Forecast Report was the first to contain RTO CP and zonal NCP forecasts produced by PJM staff. Unlike previous forecasts, there was no energy forecast for the zones or RTO published in the 2006 PJM Load Forecast Report.

PJM RTO Energy Model Development – April 2006

The load forecasting models were developed to provide input to system planning functions such as Regional Transmission Expansion Planning and the Installed Reserve Margin Study. Based on the needs of the System Planning Division, it was decided that an energy model was not required.

Subsequent to the publication of the 2006 PJM Load Forecast Report, however, several PJM corporate needs for an energy forecast were identified. Through ReliabilityFirst Corporation, PJM still had a need for an energy forecast to satisfy certain NERC reporting requirements. In addition, an energy forecast was requested for the development and monitoring of the PJM Stated Rate. To satisfy these needs, CAP staff modified the PJM RTO NCP model to produce an energy forecasting model. This modification consisted of changing the model's dependent variable from daily peak load to daily energy consumption.

The PJM RTO energy model statistics, such as goodness of fit and forecasting ability, were equivalent to similar statistics for the PJM RTO NCP model. Goodness of fit statistics are commonly used to judge how well a model represents actual energy consumption and daily RTO peak demand. Since the structure of the independent variables is the same in both models, the same economic drivers and weather simulation could be used. While the model is producing daily energy consumption, this energy can be aggregated to monthly and annual values. The PJM energy forecast will be updated on a monthly basis to coincide with the release of a new economic forecast from Moody's Economy.com.

A significant difference between the PJM RTO load model and the PJM RTO energy model is the issue of weather normalization. To evaluate the effectiveness of the NCP load model, the summer and winter peaks are weather normalized and compared to the forecast. This weather normalization is essentially being done for two hours of the year (the summer and winter peaks). The PJM RTO energy model forecasts daily energy which has a weather effect for each hour of the day. To evaluate how well the energy model is performing, it is, in effect, necessary to weather normalize 8760 hours of the year.

Preliminary results from the energy model using the historical weather simulations for a one year forecast indicate fairly tight monthly distributions about the median forecast. Current analysis on the weather drivers for the model is attempting to determine if it is these drivers that are producing a tight weather band about median weather.

The PJM energy forecasting model is still in development. While preliminary indications are positive, additional work needs to be done in several areas, including benchmarking of results, analysis of bands about the median forecast and weather normalization.

TECHNICAL DISCUSSION

The long-term daily NCP model is a linear regression model of daily NCP loads. A separate model was estimated for each PJM zone using non-coincident peak loads as the dependent variable. Although the model structure was the same for each zone, the estimation process develops a set of model coefficients specific to each zone. The model, as used for the 2006 PJM Load Forecast Report, contains the following explanatory variables:

Day-of-the-Week Binary Variables. This set of variables accounts for differences in the daily peak by day-of-the-week. The binary variables included are: Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday. Sunday is captured by the Intercept term.

Holiday Binary and Graduated Variables. This set of variables accounts for load variations associated with holidays. There is a separate binary variable for the following days: New Year's Day, Martin Luther King Day, President's Day, Good Friday, Memorial Day, July 4th, Labor Day, Thanksgiving, the Friday after Thanksgiving, Christmas Eve, Christmas Day and New Year's Eve. In addition, graduated variables are introduced to capture the reduced loads occurring during the Christmas and New Year's holiday weeks. These variables capture the fact that many schools and businesses operate differently during this time period.

Monthly Binary Variables. Monthly binary variables capture some of the seasonal variations in loads. The monthly binary variables included in the model are: January, February, March, April, May, June, July, August, September, October, November. December is captured by the Intercept term.

Solar Data. The sunrise and sunset times are entered into the model to capture the load changes associated with changes in the number of hours of daylight. They are entered into the model as separate variables, rather than constrained as a single hours-of-light variable. A daylight savings binary variable accounts for changes in load behavior during daylight savings time.

Economic Drivers. Real Gross Metropolitan Product (GMP) is used as the economic growth driver. This variable is introduced in levels and interacted with a Heating Degree Day variable and a Cooling Degree Day variable to capture the influence of more weather-sensitive equipment due to load growth. Table 1 below shows which Metropolitan Statistical Areas are assigned to each zone.

Table -1: Mapping of Load Zones to Economic Areas

Zone Name	MSA Abbreviation	Metro Area Name
AE	MATA	Atlantic City NJ
AEP	MFOW, MSOU, MKAL, MCAN, MCOU, MROA, MCHW	Fort Wayne IN, South Bend IN, Kalamazoo MI, Canton OH, Columbus OH, Roanoke VA, Charleston WV
APS	MHAS, MMOG	Hagerstown MD, Morgantown WV
BGE	MBAL	Baltimore MD
COMED	DMCHI	Chicago IL
DAY	MDAY	Dayton OH
DLCO	MPIT	Pittsburgh PA
DOM	MRIC, MVIR, MROA	Richmond VA, Virginia Beach VA, Roanoke VA
DPL	MDOV, DMWIL, MSAS	Dover DE, Wilmington DE, Salisbury MD
JCPL	DMCAM, DMEDI, MTRE	Camden NJ, Edison NJ, Trenton NJ
METED	MALL, MREA, MYOR, MLEB	Allentown/Bethlehem/Easton PA, Reading PA, York PA, Lebanon PA
PECO	DMPHI	Philadelphia PA
PENLC	MERI, MALT, MJOH	Erie PA, Altoona PA, Johnstown PA
PEPCO	DMWAS	Washington D.C.
PL	MALL, MSWB, MHAI, MLAC, MWII	Allentown/Bethlehem/Easton PA, Scranton Wilkes Barre PA, Harrisburg PA, Lancaster PA, Williamsport PA
PS	DMCAM, DMEDI, MTRE, DMNEA	Camden NJ, Edison NJ, Trenton NJ, Newark NJ
RECO	DMNEA	Newark NJ
UGI	MSWB	Scranton Wilkes Barre PA

Weather Variables. Weather is a key factor influencing PJM peaks. PJM has a rich historical database of hourly temperature, humidity, cloud cover and wind speed data. As a result, several approaches were reviewed to determine a set of weather variables that should be included in the model. Data for 17 weather stations were selected for this project. The mapping of weather stations to zones is presented below.

Table -2: Mapping of Load Zones to Weather Stations

Zone	Region	Weather Station	Call Letters
AE	PJM Middle Atlantic	Atlantic City	ACY
BGE	PJM Middle Atlantic	Baltimore-Washington International Airport	BWI
DPL	PJM Middle Atlantic	Philadelphia International Airport	PHL
JCPL	PJM Middle Atlantic	Atlantic City	ACY
		Newark International Airport	EWR
METED	PJM Middle Atlantic	Allentown Lehigh Valley International Airport	ABE
		Philadelphia International Airport	PHL
PECO	PJM Middle Atlantic	Philadelphia International Airport	PHL
PENLC	PJM Middle Atlantic	Erie International Airport	ERI
		Williamsport Regional Airport	IPT
PEPCO	PJM Middle Atlantic	Washington Reagan National Airport	DCA
PPL	PJM Middle Atlantic	Allentown Lehigh Valley International Airport	ABE
		Williamsport Regional Airport	IPT
PS	PJM Middle Atlantic	Newark International Airport	EWR
RECO	PJM Middle Atlantic	Newark International Airport	EWR
UGI	PJM Middle Atlantic	Wilkes-Barre Scranton International Airport	AVP
AEP	PJM Western	Charleston Yeager Airport	CRW
		Columbus Port Columbus International Airport	CMH
APS	PJM Western	Washington Dulles International Airport	IAD
		Pittsburgh International Airport	PIT
COMED	PJM Western	Chicago O'Hare International Airport	ORD
DAYTON	PJM Western	Columbus Port Columbus International Airport	CMH
DQE	PJM Western	Pittsburgh International Airport	PIT
DOM	PJM Southern	Richmond International Airport	RIC
		Norfolk International Airport	ORF
		Washington Dulles International Airport	IAD

THI Variables. The first step in constructing the weather variables was to define a set of hourly THI variables using the following formula:

$$\text{THI}_h = \text{Drybulb}_h - (0.55 \times (1 - \text{Humidity}_h)) \times (\text{Drybulb - 58})$$

The resulting hourly THI variables were then used to construct four time-of-day THI blocks:

- Night THI (the average of the hourly THI values from 0100 to 0500),
- Morning THI (the average of the hourly THI values from 0600 to 1100),
- Afternoon THI (the average of the hourly THI values from 1200 to 1700), and
- Evening THI (the average of the hourly THI values from 1800 to 2400).

Heating and cooling THI splines were then defined for each of these time-of-day periods, yielding eight (8) time-of-day THI variables. These variables are defined as follows:

NightHDD = Max(50 - NightTHI,0)
Morning HDD = Max(55 - MornTHI,0)
AfternoonHDD = Max(62 - AftTHI,0)
EveningHDD Max(55 - EvenTHI,0)

NightCDD = Max(NightTHI - 60,0)
MorningCDD = Max(MornTHI - 65,0)
AfternoonCDD = Max(AftTHI - 72,0)
EveningCDD = Max(EvenTHI - 65,0)

To account for differences in the weather response between weekdays and weekends, these time-of-day THI variables were augmented with the following variables.

AvgTHI = the average THI for the day
WeekendCDD = Weekend x CDD
WeekendHDD = Weekend x HDD

Where:

HDD = Max(60 - AvgTHI,0)
CDD = Max(AvgTHI - 65,0)

To reduce model complexity, summary variables were used instead of interactions of the time-of-day THI variables with the weekend binary variables. Lagged average THI heating and cooling effects were also included to capture weather buildup effects.

When estimating a model over a long data period it is important to allow for varying load response to weather, attributable to greater saturation of weather-sensitive equipment. For example, the load response to a temperature increase from 70 degrees to 75 degrees is likely to be higher in 2004 than in 1998. This reflects more homes and businesses, as well as increased saturation of cooling equipment in existing homes and businesses. To model the trend in weather response, the HDD and CDD variables (based on the Average THI value) were interacted with the GMP variable.

Although humidity is bound with temperature through the THI variable, humidity is allowed to have a separate effect by adding an average humidity variable for the day.

FuzzyHum = Min(Max(AvgTemp-60,0)/15,1)*AvgHUM

The impact of average humidity begins when the average temperature rises above 60 degrees. The impact is phased in until temperatures reach 75 degrees. At 75 degrees, average humidity is considered to have reached its full impact.

Wind speed. Wind speed impacts electrical load differently, depending on whether conditions are hot or cold. When it is hot, wind acts as a cooling agent, reducing load. When it is cold, there is a wind chill affect, increasing load. To allow the model to separate these two effects, the following two variables were introduced into the model:

ColdWind = Min(Max(60 - AfternoonTHI,0)/20,1)*AvgWSP
HotWind = Min(Max(AfternoonTHI-60,0)/15,1)*AvgWSP.

The ColdWind variable's influence begins when afternoon THI values fall below 60 degrees. The impact of wind is ramped until THI values fall below 40 degrees. Below 40 degrees wind speed has it full impact on loads.

The HotWind variable's influence begins when afternoon THI values are above 60 degrees. The impact is ramped until THI values are above 75 degrees.

Model Estimation Data Development

Zonal Daily NCP. Zonal unrestricted hourly loads were used to determine the daily zonal maximum load (NCP). The zonal unrestricted load is the metered load plus any invoked Active Load Management (ALM), voltage reduction, or significant loss of load. The zonal metered load and estimated impacts of ALM, voltage reductions and lost load are stored and maintained in the PJM Information Warehouse (PIW).

Zonal Weather Data. Historical hourly weather observations for zonal weather stations are also stored and maintained in the PIW. The hourly weather data was converted to the model weather parameters and weighted by weather station for each zone (as detailed in Table 2).

Calendar Data. Calendar and time related data was incorporated into the zonal models.

Economic Data. Historical economic data was obtained from Moody's Economy.com. In particular, the Gross Metropolitan Product (GMP) of major metros was aggregated by zone to develop zone-specific GMP (as detailed in Table 1).

Model Parameter Estimation

The model specified in E1 above was estimated for the historical period for which data was available. The beginning of the estimation period (January 1998) was determined by the date for which metered hourly loads were available for all PJM zones. The end of the estimation period was determined by the availability of current load, weather and economic data.

The model development and estimation was done using the MetrixND model-building software from Itron, Inc. MetrixND allows for easy definition of additional weather parameters such as heating degree days and cooling degree days (HDD, CDD) from the THI parameter. It also enables variable interaction in the case of HDD and CDD with GMP. The calendar binary variable day-of-the-week provides relative NCP magnitude differences within a week and the calendar binary variable month provides differences within the year.

Once the model coefficients are estimated, PJM staff analyzes and evaluates the common statistical measures of goodness of fit. In addition, the model is checked to see how well it performed in predicting zonal and RTO historical peak days (backcasting).

Forecast Model: Forecast Data Requirements

The Forecast Model structure described in E1 requires forecast period projections for the temporal, economic and weather drivers. The temporal calendar and weather projections, such as day-of-week, month, holiday, leap year, DST, sunrise and sunset are known. The economic data projections use economic forecast scenarios obtained from Moody's Economy .Com. The weather projections are obtained by using historical daily weather in a Monte Carlo simulation.

Monte Carlo Simulation

The Forecast Model E1 is solved by moving through the year, day by day, using forecasted economic variables and historical weather patterns for each day. The economic scenario used for the base forecast is the "most likely" economic scenario created by Moody's Economy.Com.

To model the most likely weather conditions (often referred to as the expected or normal weather) a Monte Carlo technique is used to simulate a distribution of daily load scenarios generated by historical weather observations, representing actual weather patterns that prevailed across the PJM control region.

To enhance the simulation process, each yearly weather pattern is shifted by each day of the week moving forward, providing 7 different weather scenarios for each historical year. For late December dates, data from January of the same calendar year is applied. Table 3 below illustrates the shift of weather data across the scenarios.

Table 3: Mapping of Weather Scenarios to Dates

	Scenario						
Date	A1971	B1971	C1971	D1971	E1971	F1971	G1971
1-Jan	1/1/1971	1/2/1971	1/3/1971	1/4/1971	1/5/1971	1/6/1971	1/7/1971
2-Jan	1/2/1971	1/3/1971	1/4/1971	1/5/1971	1/6/1971	1/7/1971	1/8/1971
3-Jan	1/3/1971	1/4/1971	1/5/1971	1/6/1971	1/7/1971	1/8/1971	1/9/1971
4-Jan	1/4/1971	1/5/1971	1/6/1971	1/7/1971	1/8/1971	1/9/1971	1/10/1971
5-Jan	1/5/1971	1/6/1971	1/7/1971	1/8/1971	1/9/1971	1/10/1971	1/11/1971
6-Jan	1/6/1971	1/7/1971	1/8/1971	1/9/1971	1/10/1971	1/11/1971	1/12/1971
7-Jan	1/7/1971	1/8/1971	1/9/1971	1/10/1971	1/11/1971	1/12/1971	1/13/1971
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
25-Dec	12/25/1971	12/26/1971	12/27/1971	12/28/1971	12/29/1971	12/30/1971	12/31/1971
26-Dec	12/26/1971	12/27/1971	12/28/1971	12/29/1971	12/30/1971	12/31/1971	1/1/1971
27-Dec	12/27/1971	12/28/1971	12/29/1971	12/30/1971	12/31/1971	1/1/1971	1/2/1971
28-Dec	12/28/1971	12/29/1971	12/30/1971	12/31/1971	1/1/1971	1/2/1971	1/3/1971
29-Dec	12/29/1971	12/30/1971	12/31/1971	1/1/1971	1/2/1971	1/3/1971	1/4/1971
30-Dec	12/30/1971	12/31/1971	1/1/1971	1/2/1971	1/3/1971	1/4/1971	1/5/1971
31-Dec	12/31/1971	1/1/1971	1/2/1971	1/3/1971	1/4/1971	1/5/1971	1/6/1971

This approach has two key advantages. One, by rotating the data on the calendar, peak producing weather occurs on peak producing days. Two, by lining up the actual weather patterns across all weather stations for a day (say, January 5, 1984) the diversity of weather patterns that impact the PJM total system is simulated. This is a more plausible approach to peak forecasting, as opposed to traditional methods that would have all weather stations having peak producing weather on the same day. The latter approach would overstate the PJM total peak forecast by minimizing diversity.

The process is repeated for the remaining years of historical weather data for each year in the forecast horizon. These simulations produce a frequency distribution of NCP demands by zone. The NCP values are summed across the zones to provide a forecast of the System NCP.

As stated, the simulation process produces a distribution of forecasts of daily NCPs (by zone) from the years of historical weather. While the forecasting process has the capability of producing daily NCP forecasts, for purposes of the Load Report which provides data to support reliability planning, only monthly and annual NCP forecasts by zones were retained. For each weather scenario, monthly NCPs are determined by obtaining the maximum NCP for the month. Annual NCPs are determined by the maximum over the 12 months.

In addition to the zonal NCP forecasts, an RTO NCP forecast equal to the sum of the zonal NCPs is produced.

Diversity and Coincident Peaks

Of all the issues encountered in developing the PJM Forecast Modeling Process, a forecast of the Zonal CPs was the greatest challenge. In the original approach developed by Itron, independent models for the NCP and CP were utilized. Since these models were independent, the resulting forecasts did not necessarily maintain the historical diversities of the zones. This was demonstrated by results which, in some years of the forecast and for some zones, the CP forecast exceeded the NCP forecast.

Itron modified the approach by using the NCP and CP models to simulate 2002 for each zone and each weather scenario. The year 2002 was selected because it fell roughly in the middle of the historical period over which the NCP and CP models were estimated. For each zone, the maximum load value from the daily NCP simulations were computed for each month. This step resulted in 12 NCP values for each of the historical weather scenarios. For each weather scenario, the daily CP projections were summed across the zones. This formed an estimate of the System CP for each day. Within each month, the maximum RTO CP value was identified. The zone CP values that corresponded to the day of the System CP maximum were then recorded. This step resulted in 12 CP values for each of the historical weather scenarios. For each weather scenario, the monthly diversity factors were computed as the ratio of the monthly CP value to the monthly NCP value. This step resulted in 12 diversity values for each of the historical weather scenarios. To ensure that the simulated diversity factors were consistent with observed diversity factors over the period 1998 through 2004, the frequency distributions of actual versus simulated diversity factors were compared for each zone.

PJM was concerned with only using a one-year interval to estimate diversity. Conceptually, PJM thought this may not incorporate enough system loads to accurately reflect a range of diversity. Summary statistics calculated from the diversity factors provided by Itron seemed to understate historical average RTO diversity. Ideally, PJM wanted to model diversity to coincide with historically observed diversity. PJM decided to estimate diversity using a longer time interval, using three years (2000-2002) instead of one year (2001). Using three years to estimate the diversity, PJM found that the forecasts of NCP and CP again crossed over, implying negative diversity factors. The problems diversity was causing between the NCP and CP forecasts which were identified during the initial development stage with Itron persisted. Upon further investigation, it was determined that the crossover issue resulted from the zonal NCP share of the NCP total being less than the zonal CP share. This increased CP share of the RTO system total caused the CP to be greater than the NCP. The relationship of a zone's share of the RTO was overriding the effect of diversity.

After many false starts, a solution to reflect consistency between the NCP and the CP forecasts' diversities was developed. The procedure for developing consistent diversities is:

- **Overview of Developing Zonal Diversities:** To simulate RTO diversity, the PJM Load Forecast process first estimates the monthly NCP per zone per weather scenario. The zonal CP is calculated by multiplying the monthly zonal NCP by the monthly (estimated) zonal diversity that would occur under those same weather conditions. The remainder of the process is devoted to estimating the zonal diversity factor the same weather conditions would produce. The available historical zonal load data is limited and would produce for the actual weather conditions over the last eight years a zonal diversity factor for each month of the year. A distribution of diversity factors for each zone and month is created using the Crystal Ball software that employs a Monte Carlo simulation to produce distributions for defined parameters. These distributions must be coupled with the historical weather scenarios. This is achieved by developing a distribution of monthly zonal CP share of the PJM RTO peak. The distribution of diversity factors and distribution of monthly zonal CP shares are convolved to obtain a distribution of diversity factors by weather scenario.

- **Estimating Zonal Diversity Factors Using Crystal Ball:** The Crystal Ball software is used to develop a historical distribution of diversity factors for each zone and each month of the year. Using the eight years of historical zonal loads, monthly zonal diversity factors are computed. The eight years of historical diversity factors produce summary statistics which are used in Crystal Ball to define a distribution for each zone and each month of the year. Crystal Ball then simulates a distribution of diversity factors using a Monte Carlo technique. Crystal Ball produces a distribution of diversity factors corresponding to the historical weather scenarios.

- **Modeling of Zonal CP Shares of RTO Peak:** The zonal CP share of the daily RTO peak is modeled using a neural network model. The general model structure uses day of week and an intercept in the linear node. The two non-linear nodes incorporate weather and economic concepts. Three nodes were chosen based upon visual plots and fit statistics (using four and five nodes resulted in only minor mathematical improvement while a visual improvement seemed minimal). It is believed the three nodes provide a more generic relationship of share vs. weather rather than purely maximizing in-sample fit. The estimation period was 1/1/1998 through 10/31/2005. The model is solved over the forecast period and for all historical weather simulations. The zonal shares of the daily RTO peak are averaged to obtain the zonal monthly share per month and weather scenario.

- **Calculating Monthly Zonal CP by Weather Scenario:** The sum of zonal NCP forecasts by month/scenario creates a RTO "Sum of NCP" value, which is then adjusted by historical monthly average RTO diversity to create a RTO system forecast by month/scenario. Zonal forecasted CP share of RTO peak (from the neural net models) is joined with the zonal monthly NCP forecasts. The RTO System forecast multiplied by the zonal CP share by month/scenario creates a CP forecast for each zone/month/scenario combination.

- **<u>Produce Forecast of Zonal NCP and CP by Month/Scenario:</u>** The forecasted zonal NCP and CP create a forecasted diversity by month/scenario. Each distribution by zone/month/scenario is sorted by forecasted diversity. Each of these distributions is combined with the Crystal Ball produced distribution of historical diversities sorted in the same manner as the forecasted diversities. This yields forecasts of zonal NCP and CP by zone/month/scenario that have consistent diversity.

ONGOING MODEL ENHANCEMENTS/ INDEPENDENT REVIEW

Subsequent to the release of the 2006 PJM Load Forecast Report, PJM initiated a series of enhancements to both the forecast model and its processing. A number of these changes were identified during the preparation of the 2006 forecast but could not be implemented due to time constraints. Others were identified as a result of post-forecast analysis. Following is a list of implemented enhancements and their impact:

- **Solving the Model per Weather Scenario:** Solving the model for each weather scenarios had been done in the Access application, which required processing time, for a forecast horizon of 15 years, of 48 to 72 hours. Converting this process to SAS software reduced the processing time to 1 to 2 hours;

- **Number of Weather Scenarios:** The 2006 forecast produced 7 weather scenarios per day for each historic weather year. This was increased to 13 scenarios per day for each historic weather year, which improved the distribution of load response to weather and better defines normal and extreme weather response;

- **Weather Variables:** The single daily Average Temperature-Humidity Index variable was separated into seasonal variables: Maximum Temperature-Humidity Index for the cooling season, Minimum Wind Speed Adjusted Temperature for the heating season, and Average Temperature for shoulder months. This change improved the model's ability to distinguish weather response between seasons;

- **Length of Day:** Separate variables for Sunrise and Sunset were replaced with a single measure, minutes of daylight. This resulted in modest improvement in model performance.

- **Seasonal Peaks:** The processing of model results now identifies each zone's peak over a range of months as the seasonal peak. Previously, January and July peaks were designated as seasonal peaks. This raised zonal seasonal peaks by up to 2%, and also resulted in smooth growth patterns for zones that previously displayed a variable year-to-year growth pattern;

- **Diversity:** Diversity was modeled, replacing the use of average observed historical diversity. This had the beneficial impact that diversity now varies across weather scenarios, from relatively larger diversity at mild weather to little to no diversity at extreme weather.

In addition to these changes which were implemented for the 2007 PJM Load Forecast Report, PJM has received recommendations from an outside consultant (The Brattle Group) who was engaged in Fall 2006 to provide an independent review of the forecast model and process developed by PJM staff. Those recommendations include:

- **Transformed Dependent Variable:** Transforming the dependent variable into natural logs will minimize any heteroscedasticity that may exist and may also improve forecasting performance;

- **Autocorrelation Correction:** Zonal models exhibit high correlation of error terms, which can be rectified by adding a first-order autoregressive term to the model;

- **Parsimony:** Independent variables which are not contributing appreciably to forecast performance should be removed from the model specification;

- **Separate Seasonal Models:** Estimates of weather response may improve further if separate models are fitted for the seasons or months, since the weather coefficients vary appreciably from period to period;

- **Joint Estimation Across Zones:** Joint estimation of the 18 zonal models, using Zellner's Seemingly Unrelated Regression, may improve the efficiency of estimation. This technique would allow for estimated parameters for each zone to be influenced by those of other zones.

This document will be updated as the model continues to be revised.